Structural basis for a six nucleotide genetic alphabet

Millie M. Georgiadis,[‡, †,*] Isha Singh,[‡] Whitney F. Kellett,[†] Shuichi Hoshika,[§] Steven A. Benner,[§*] and Nigel G. J. Richards[†]

[†]Department of Chemistry & Chemical Biology, Indiana University Purdue University Indianapolis, Indianapolis, IN 46202, USA

[‡]Department of Biochemistry & Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

[§]Foundation for Applied Molecular Evolution and the Westheimer Institute of Science & Technology, Gainesville, FL 32601, USA

[*]To whom correspondence should be addressed: M.M.G.: telephone, (317) 278-8486; fax, (317) 274-4686; e-mail, mgeorgia@iu.edu.

This is the author's manuscript of the article published in final edited form as:
Georgiadis, M. M., Singh, I., Kellett, W. F., Hoshika, S., Benner, S. A., & Richards, N. G. J. (2015). Structural basis for a six nucleotide genetic alphabet. Journal of the American Chemical Society, 137(21), 6947-6955.
http://dx.doi.org/10.1021/jacs.5b03482

ABSTRACT

Expanded genetic systems are most likely to work with natural enzymes if the added nucleotides pair with geometries that are similar to those displayed by standard duplex DNA. Here, we present crystal structures of 16-mer duplexes showing this to be the case with two non-standard nucleobases (**Z**, 6-amino-5-nitro-2(1H)-pyridone and **P**, 2-amino-imidazo[1,2-a]-1,3,5-triazin-4(8H)one) that were designed to form a **Z**:**P** pair with a standard "edge on" Watson-Crick geometry, but joined by rearranged hydrogen bond donor and acceptor groups. One duplex, with four **Z**:**P** pairs, was crystallized with a reverse transcriptase host and adopts primarily a B-form. Another contained *six* consecutive **Z**:**P** pairs; it crystallized without a host in an A-form. In both structures, **Z**:**P** pairs fit canonical nucleobase hydrogen-bonding parameters and known DNA helical forms. Unique features include stacking of the nitro group on **Z** with the adjacent nucleobase ring in the A-form duplex. In both B-and A-duplexes, major groove widths for the **Z**:**P** pairs are approximately 1 Å wider than those of comparable G:C pairs, perhaps to accommodate the large nitro group on **Z**. Otherwise, **ZP**-rich DNA had many of the same properties as CG-rich DNA, a conclusion supported by circular dichroism studies in solution. The ability of standard duplexes to accommodate multiple and consecutive **Z**:**P** pairs is consistent with the ability of natural polymerases to biosynthesize those pairs. This, in turn, implies that the GACT**ZP** synthetic genetic system can explore the entire expanded sequence space that additional nucleotides create, a major step forward in this area of synthetic biology.

KEYWORDS: Non-natural base pairs; DNA structure; host-guest complex; synthetic biology, A-DNA; B-DNA

INTRODUCTION

Originally defined as simply the field seeking to create artificial life,[1] "synthetic biology" became more narrowly applied in the 1970s to mean the use of emerging recombinant DNA technologies to rearrange and modify *natural* genetic pieces.[2] Soon, however, synthetic biologists will no longer be constrained to exploiting or manipulating the natural nucleotides and amino acids found in modern terran biopolymers. Instead, several kinds of nucleotide analogues have been reported to form non-standard nucleobase pairs "orthogonal" to the standard T:A and C:G pairs when they are incorporated into DNA.[3-14] Further, in many cases, these extra deoxyribonucleotides can direct the synthesis of non-standard RNA containing extra ribonucleotides, thereby increasing the number of codons that, in turn, can encode proteins containing additional amino acids.[15-18] These developments have led to a new view of synthetic biology, adumbrated by Eric Kool,[9] which seeks to create properties that we value in life (including reproduction, adaptation, and evolution) using molecular platforms that combine natural and synthetic components.

In some cases, however, molecular behaviors peculiar to the added synthetic genetic components limit that combination. For example, Romesberg has shown that two nonstandard nucleobases, d5SICS and dNaM[19-22] designed to pair by hydrophobic and geometric complementarity, exhibit greater than 95% retention in duplex DNA after 15 hours of growth in living *Escherichia coli* cells.[23] However, a potentially undesirable property of d5SICS:dNaM within duplex DNA is its intrinsic pairing *not* in the designed "edge-on" geometry, but rather by inter-strand "stacking" of the two hydrophobic species. This stacking introduces significant distortion of the duplex DNA backbone.[24] Only when bound within the active site of a polymerase is the d5SICS:dNaM pair forced into the Watson-Crick pairing geometry.[25] One might therefore predict that inclusion of many d5SICS:dNaM pairs at multiple or

consecutive sites would compromise overall DNA conformation, especially in light of the toxicity of intercalative DNA binding agents, which distort the DNA structure and prevent protein binding with deleterious consequences for gene regulation and subsequent transcriptional events.

Members of another class of artificially expanded genetic information systems (AEGIS) are neither so hydrophobic nor so geometrically different from natural nucleotides as d5SICS and dNaM.[13] Rather than pairing via hydrophobic interactions, these retain the hydrogen bonding that joins the standard A:T and G:C pairs. The pairing is "orthogonal" to the pairing of standard bases because the hydrogen bonding units are shuffled. For example, the small 6-amino-5-nitro-2(1H)-pyridone heterocycle (trivially named **Z**) presents a donor-donor-acceptor pattern of hydrogen bonding units to a complementary large 2-amino-imidazo[1,2-a]-1,3,5-triazin-4(8H)one heterocycle (trivially named **P**), which presents an acceptor-acceptor-donor pattern of hydrogen bonding units (Figure 1).[26] Further, **Z** brings to nucleic acids a functional moiety (the nitro group) that is not found in any natural encoded biopolymer, a functional group that may contribute to the ability of GACT**ZP** DNA to deliver especially effective binders via *in vitro* selection experiments.[27]

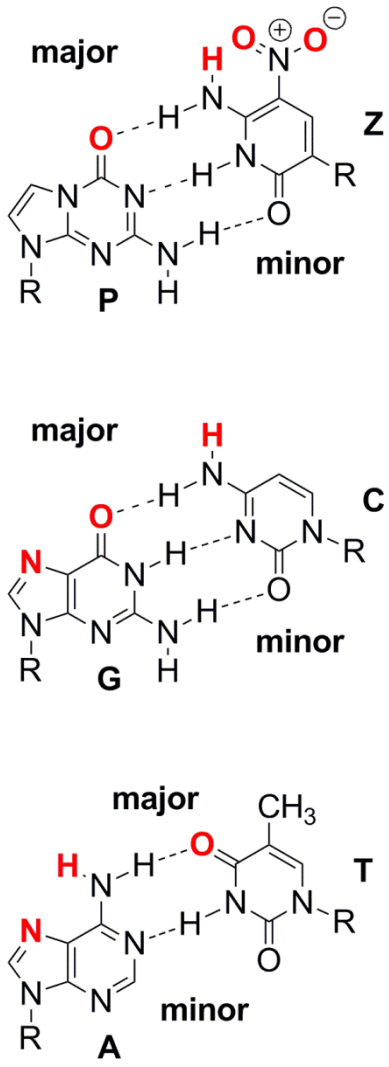

Figure 1. Chemical structures of **Z**:**P**, G:C, and A:T paired nucleobases with atoms that can hydrogen bond in the major or minor groove highlighted in red text.

Polymerase experiments with the **Z**:**P** pair and other AEGIS pairs have been auspicious. DNA containing up to four consecutive **Z** or **P** nucleotides can be PCR amplified by polymerases.[28,29] However, as learned from studies in unnatural nucleotides that pair without hydrogen bonding,[25] this does not mean that absent the constraints of a polymerase active site, the **Z**:**P** has a canonical geometry. Further, while X-ray crystallography has in the past been used to examine DNA containing mismatches to AEGIS components,[30] no crystallographic studies have yet provided information for the geometry of

correctly matched AEGIS pairs, or about the overall impact upon duplex structure of DNA that multiple non-natural nucleobase pairs might have, even for the d5SICS:dNaM system. Here, we provide these data.

EXPERIMENTAL PROCEDURES

Synthesis and purification of Z-P containing oligonucleotides

Standard phosphoramidites (Bz-dA, Ac-dC, dmf-dG, dT, and 5-Br-dU-CE) and controlled pore glass (CPG) having standard nucleosides were purchased from Glen Research (Sterling, VA). AEGIS phosphoramidites (d**Z** and d**P**) were obtained from Firebird Biomolecular Sciences, LLC (www.firebirdbio.com, Alachua, FL). All oligonucleotides containing **Z** and **P** (2P, 3/6ZP, 3/6ZP Br1 and 3/6ZP Br2, see Figure 1 and Table 1) were synthesized on an ABI 394 DNA Synthesizer following standard phosphoramidite chemistry, as previously reported.[30] The CPGs carrying the synthetic oligonucleotides were treated with 1 M DBU in anhydrous acetonitrile (2.0 mL) at room temperature for 24 hours to remove the NPE group from the **Z** nucleobase. Then, the CPGs were filtered and dried. The CPGs carrying 2P and 3/6ZP (without 5-Br-dU) were treated with concentrated ammonium hydroxide at 55 ºC for 16 hours, while the CPGs carrying 3/6ZP Br1 and 3/6ZP Br2 (with 5-Br-dU) were treated with concentrated ammonium hydroxide at room temperature for 24 hours. After removal of ammonium hydroxide, the oligonucleotides containing **Z** and **P** were purified on ion-exchange HPLC, and then desalted using Sep-Pac® Plus C18 cartridges (Waters). Fully standard oligonucleotides were purchased from Midland Certified Reagent Co. (Midland, Texas) in desalted form and used without further purification.

Table 1. Oligonucleotides used in this study

Name	# adjacentZPs/ total	Sequence

	ZPs	
2P	2/4	5′-CTTATPPTAZZATAAG
3/6ZP	6/6	5′-CTTATPPPZZZATAAG
3/6 ZP Br1	6/6	5′-CT**B**ATPPPZZZATAAG
3/6 ZP Br2	6/6	5′-CT**B**ATPPPZZZA**B**AAG
AT		5′-CTTATAAATTTATAAG
GC		5′-CTTATGGGCCCATAAG

B indicates replacement of thymine with 5-bromouracil.

Circular dichroism experiments

For CD analysis, the 2.5 mM stocks of 3/6 ZP (5'- CTTAT**PPPZZZ**ATAAG -3'), AT sequence (5' CTTATAAATTTATAAG 3'), GC sequence (5' CTTATGGGCCCATAAG 3'), and 2P (5' CTTATPPTAZZATAAG) were diluted to 5 µM in 10 mM HEPES pH 7.0 and 10 mM $MgCl_2$, the buffer used to prepare the oligonucleotides for crystallization. The CD spectra for DNA sequences were collected on a Jasco J-810 CD instrument at 25 °C, at a rate of 50 nm/min and a wavelength increment of 0.1 nm. Ellipticity, Ø (mdegrees) was recorded for the DNA sequences from a wavelength of 320 to 220 nm. The spectra were the average of five scans corrected for ellipticity readings obtained for buffer (10 mM HEPES pH 7.0, 10 mM $MgCl_2$) by itself. Spectra were initially measured for 3/6 ZP, GC and AT control sequences and subsequently for 2P, GC, and AT control sequences.

Crystallization of ZP containing and other oligonucleotides

Sixteen nucleotide self-complementary oligonucleotides containing two or six consecutive **Z**:**P** nucleobase pairs were designed to be compatible with our host-guest crystallization. The host in this system is the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase, purified as previously described;[31] the guests are the various oligonucleotide duplexes. Oligonucleotide sequences screened for crystallization are compiled in Table 1. The oligonucleotides were resuspended in buffer containing 10 mM HEPES pH 7.0 and 10 mM $MgCl_2$ to give a final concentration of 2.5 mM duplex DNA and then annealed by heating to 70 ºC followed by slow cooling to room temperature prior to crystallization. A 2.9 mM stock solution of the protein was diluted to 1.4 mM using 50 mM MES pH 6.0 and 0.3 M NaCl. This 1.4 mM sub-stock was then further diluted to 0.65 mM in 100 mM HEPES pH 7.5 and 0.3 M NaCl.

Each DNA duplex was pre-complexed with the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase with final concentrations of 0.71 and 0.46 mM, respectively, as previously described, and then subjected to self-nucleation or microseeding experiments. For microseeding, vapor diffusion hanging drops included 1 µL each of protein-DNA microseeds obtained from crystals of a host-guest complex with an ATCG DNA sequence diluted in reservoir solution containing 7 % PEG 4000, 5 mM magnesium acetate and 50 mM N-(2-acetamido)iminodiacetic acid (ADA) at pH 6.5 and 1 µL of the protein-DNA complex solution, suspended over the reservoir solution. Using a similar strategy to that described above, self-complementary 16-mer duplex DNA oligonucleotides including either G:C or A:T pairs replacing the **Z**:**P** pairs found in the 3/6ZP oligonucleotide (Table 1) were crystallized as host-guest complexes.

The duplex including 6 consecutive **Z**:**P** pairs (3/6 Z:P) did not crystallize under our normal host-guest complex conditions. However, they did crystallize in 10 mM magnesium acetate, 50 mM MES pH 5.6, and 2.5 M ammonium sulfate with the oligonucleotide at 0.35 mM. Two different oligonucleotides including 5-bromouracil replacing Ts in the sequence were synthesized (Table 1) for

phasing purposes. Both of these duplexes crystallized under similar conditions that included 10 mM magnesium acetate, 50 mM MES pH 5.6, and 1.7-2.0 M ammonium sulfate. The crystal used for the structure determination was that of the Br1 sequence, which crystallized in 10 mM magnesium acetate, 50 mM MES pH 5.6, and 1.7 M ammonium sulfate and was cryo-cooled in a solution including the reservoir with 20% glycerol added. Thus, fortuitously, we obtained crystals of oligonucleotides containing Z:P pairs both with a host and without a host, allowing the comparison of the Z:P pairs with two very different crystallization constraints.

Data collection, structure determination, and crystallographic refinement

A Br single wavelength anomalous dispersion experiment was performed to obtain phases for the 3/6 ZP-Br structure. Approximately 10-fold redundant data were collected to 1.98 Å at the APS SBC 19-ID beamline (Table 2) for a crystal of the Br1 sequence including two Br atoms from 5-bromouracil replacing thymines in the sequence as shown in Table 1. Initial Patterson searching and phasing calculations were done by using HKL3000.[32] Specifically, a single Br site was identified in SHELXD[33] and refined in SHELXE[34]. Initial phases were then calculated to 2.45 Å using MLPHARE, figure of merit 0.56, and improved and extended to 2.25 Å by solvent flattening using DM.[35-37]

This initial experimentally phased electron density map was of high quality and allowed automated building of the DNA structure by NAUTILUS,[38] albeit with A:T pairs modeled in place of **Z**:**P** pairs. Following initial refinement of the structure generated by NAUTILUS in REFMAC5,[39] positive peaks were observed in the F_o-F_c electron density maps consistent with missing chemical moieties present within the **Z** and **P** nucleotides. While maintaining the coordinate system, additional functional groups were added to the core NAUTILUS generated structure using the Molefacture Plug-in for Visual Molecular Dynamics (VMD) Graphics Viewer.[40] Upon generation of the complete coordinates, each **Z** or **P** nucleobase was optimized for ten steps using GAMESS *ab initio* molecular

quantum chemistry,[41] with a 6-31G[42] basis set. Each idealized nucleobase analog was then placed back into the model to form the completed DNA structure, including the novel functionality of the Z and P bases. Parameter files and linking statements were created for refinement in PHENIX.[43] Addition of solvent molecules and manual model adjustment was done iteratively in COOT[44] followed by refinement in PHENIX with maximum likelihood targets and isotropic B-factors. Addition of solvent molecules and manual model adjustment was done iteratively in COOT followed by refinement in PHENIX with maximum likelihood targets and isotropic B-factors.

Table 2: Crystallographic Data

Dataset	**2P-HG**	**AT-HG**	**GC-HG**	**3/6 ZP-Br**
Data statistics				
a (Å)	54.638	54.636	54.623	42.019
b (Å)	145.359	145.27	145.38	42.019
c (Å)	46.878	46.802	46.801	140.472
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P3_221$
Wavelength (Å)	0.97911	0.97933	0.97933	0.91963
Resolution (Å)	30.26-1.8	29.05-1.68	28.71-1.78	50-1.98
Total observations	199682	183063	184656	105588
Unique reflections	41412	43118	36612	10571
Completeness (%)	99.8 (96.7)	99.3 (99.3)	99.9 (100)	99.5 (98.8)

R_{merge} (%)	5.4 (42.0)	3.0 (43.1)	4.0 (38.8)	6.9 (76.8)
R_{pim}	2.6 (23.5)	1.7 (23.0)	2.0 (19.2)	2.2 (33.9)
I/σ	24.3 (3.4)	24.8 (3.4)	22.7 (4.3)	29.5 (2.3)
Refinement statistics				
R value (%)	22.0	21.6	21.6	21.2
R free (%)	24.0	23.9	23.7	23.8
RMSD bonds (Å)	0.008	0.006	0.006	0.003
RMSD angles (°)	1.168	1.115	1.064	0.829
Atoms				
Protein/DNA	2019/331	1982/325	1968/325	668
water	201	195	195	76
Average B-factors				
Protein/DNA	29.51/54.70	28.05/45.38	28.68/56.39	23.22/22.20*
water	31.96	30.09	31.08	29.0

2P, AT, GC, and 3/6 ZP refer to the DNA sequences (Table 1), HG designates a host-guest complex. Values in parentheses are for the highest resolution shell of the data: 2P-HG (1.83-1.80 Å), AT-HG (1.71-1.68 Å), GC-HG (1.82-1.78 Å), and 3/6 ZP (2.01-1.98 Å). * B-factors shown are for the A/B chains of 3/6 ZP rather than protein/DNA.

Data were collected for the 2P host-guest complex crystal at the APS SBC 19-BM beamline to Bragg spacings of 1.8 Å and for the GC and AT host-guest complex crystals at the APS GM/CA 23-ID-

D beamline to Bragg spacings of 1.78 Å and 1.7 Å, respectively (Table 2). The host-guest crystal structures were determined by molecular replacement as implemented in PHASER[45] within the CCP4 suite of programs[46] using the model of the N-terminal fragment of MMLV RT as the search model. This approach provides unbiased electron density for the DNA complexed to the protein. The protein model and associated water molecules were first adjusted in COOT and then refined initially in REFMAC and later in PHENIX to improve the electron density, and the DNA was subsequently modeled into F_o-F_c electron density maps. In building the DNA model, the first three nucleobase pairs were modeled and refined, then the next two pairs, and finally the remaining three pairs.

For the 2P structure, the **Z**:**P** nucleobase pairs were initially modeled as G:C pairs and then subsequently replaced by superimposing the base coordinates for either **Z** or **P** generated for the 3/6 ZP structure on the common atoms manually in COOT. The parameter and linking files were created for refinement in PHENIX. As the asymmetric unit includes one protein molecule and half of the DNA molecule, the DNA can be modeled as 8 pairs of duplex DNA or a single 16-mer strand of DNA. To ensure that the phosphodiester bond between bases 8 and 9 is appropriately connected, the DNA was modeled as a single 16-mer strand in the final round of the refinement, and the 8-mer duplex was then regenerated by symmetry.

Coordinates have been deposited for the 2P, GC, AT, and 3/6 ZP structures (Table 1) with PDB identifiers, 4XO0, 4XPE, 4XPC, and 4XNO, respectively.

RESULTS AND DISCUSSION

Genetic alphabets that combine standard and non-natural nucleotides will be an essential part of synthetic organisms that synthetic biologists develop in the future. The need to maintain (or at least prevent radical divergence from) existing canonical duplex structures, however, will remain an important constraint upon the incorporation of synthetic nucleotides into DNA if existing nucleic acid

binding proteins or enzymes are to interact with them. Of course, even natural nucleotides form duplexes with a considerable diversity of structures, and the special properties of stacked A:T versus G:C nucleobase pairs in duplex DNA are well documented.[47] Thus, characterizing the structural properties of DNA containing multiple and/or contiguous non-natural nucleobase pairs as compared to natural nucleobase pairs has considerable importance.

Z:P containing duplex oligonucleotides adopt right-handed B-form in solution

CD spectroscopy was used to determine the helical form of the duplexes in solution. All of the sequences (2P, AT, GC and 3/6 ZP) exhibit right-handed B-like spectra as shown in Figure 2, with some differences in the peak positions and heights pertaining to differences in the primary sequence of DNA. The spectra for the AT sequence exhibits a negative peak at 248 nm and a positive long wavelength peak at about 279 nm, typical of right handed B-DNA. The GC sequence on the other hand has a broad negative peak centered at approximately 245 nm. The positive peak for the GC sequence shifts to 270 nm instead of 279 nm observed for the AT sequence. The CD spectra of the 2P and 3/6 ZP sequences are similar to those of the GC sequence, with broad negative peaks at 241 nm and positive peaks at 273 nm. An additional feature found in the ZP-containing oligonucleotides is a second positive peak at approximately 305 nm, which may reflect specific properties of the Z nucleobase.

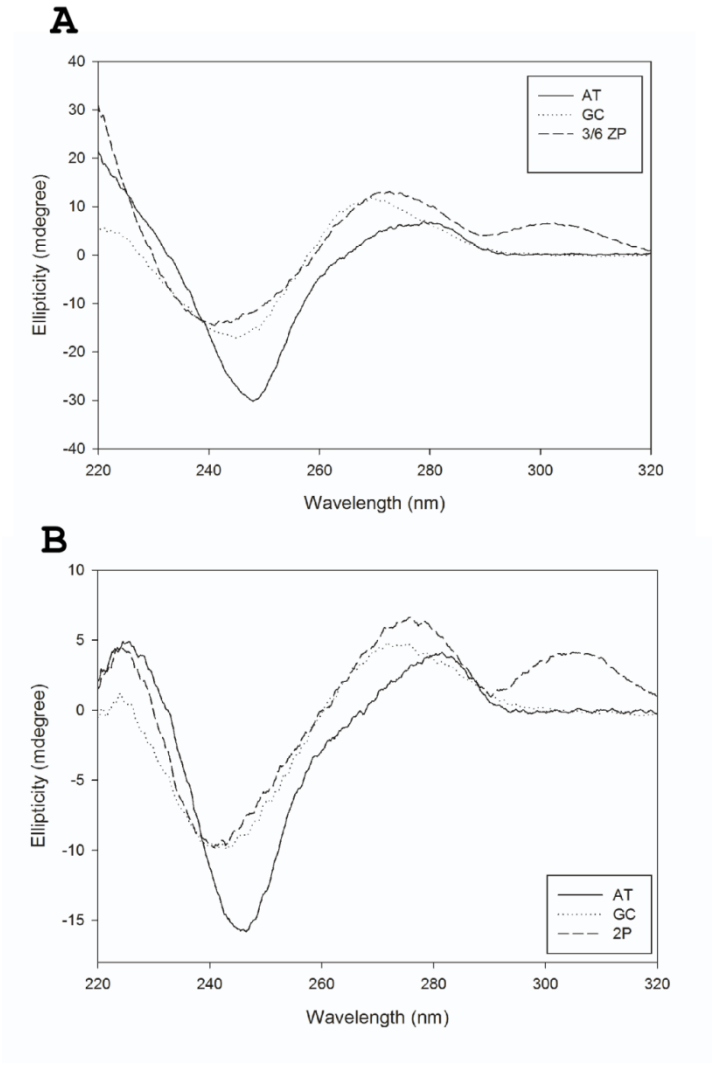

Figure 2. Characterization of oligonucleotide duplexes by circular dichroism (CD). The ellipticity is plotted versus the wavelength for (A) 3/6 ZP (dotted line) or (B) 2P (dotted line) along with control sequences for GC (dashed line), and AT (solid line). All of the duplexes have CD spectra indicative of right-handed B-form DNA.

Z:P nucleobase pairs crystallize in both host-guest and DNA only lattices

To obtain a more detailed understanding of the structural properties of **Z**:**P** nucleobase pairs, we determined the crystal structures of 2P, which includes two consecutive **Z:P** pairs, four total **Z:P** pairs, and 3/6 ZP containing six consecutive **Z:P** pairs. Our initial approach involved the use of a host-guest

system to crystallize and analyze novel nucleic acid structures. The host-guest system allows crystallization of self-complementary 16-mer oligonucleotide duplexes (guests) through complexation with a host protein, the N-terminal fragment of Moloney murine leukemia virus reverse transcriptase (MMLV-RT), which is bound to each end of the duplex. Thus, the complex contains two protein molecules bound to one 16-mer duplex (Figure 3).

The DNA guests generally exhibit B-form but can deviate significantly from this form as seen in the structures of the spore product[48] and the complex with bleomycin bound to the DNA.[49] In the host-guest complex, which also forms in solution with the same stoichiometry,[50] the host protein interacts with the three terminal nucleobase pairs on either end through minor groove hydrogen-bonding while the central 10 nucleobase pairs are free of interactions with either protein or other DNA molecules. A significant advantage of the host-guest system is that it allows for comparison of structural properties for different 16-mer oligonucleotides all within the same lattice. For example, we demonstrated that independent of its position within the oligonucleotide, a CA dinucleotide step had the same structural properties.[51] The oligonucleotide including two consecutive **Z**:**P** nucleobase pairs (2P) crystallized in the host-guest system. This structure was phased by molecular replacement using the protein structure as the search model and refined to 1.8 Å resolution.

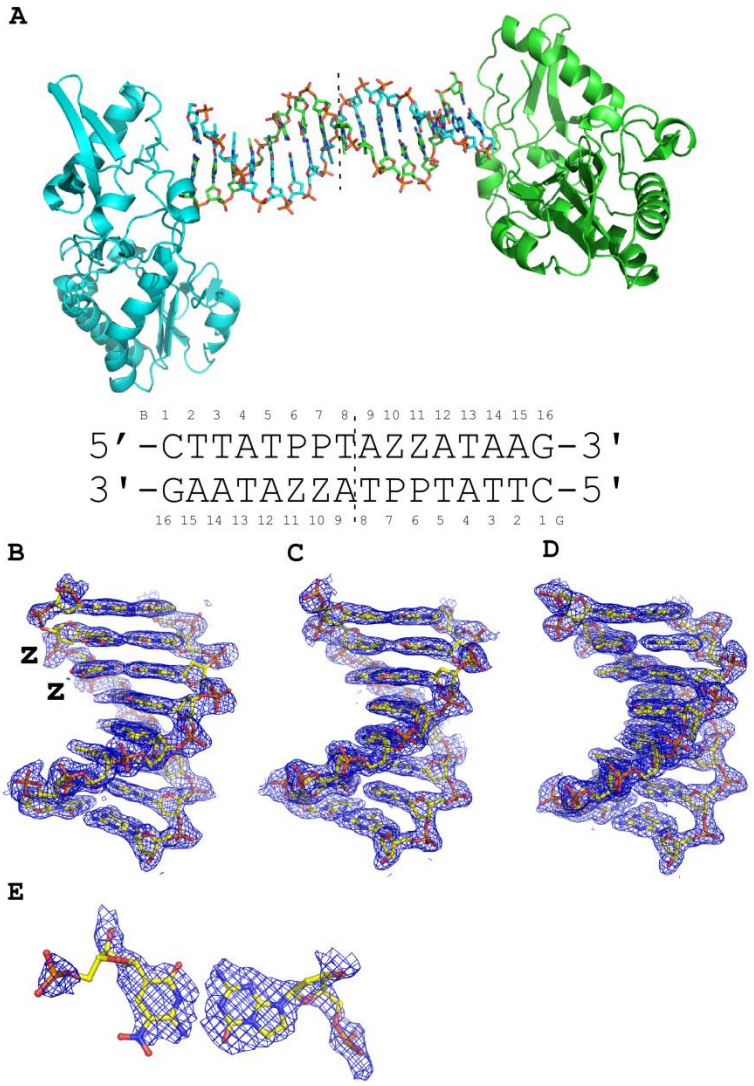

Figure 3. Crystal structures of host-guest complexes including self-complementary 16 base pair oligonucleotides. (A) The host guest complex includes two protein molecules, shown as cartoon renderings in cyan and green, and a 16-mer duplex, each strand shown as a stick rendering C, cyan or green, O, red, N, blue, and phosphorous in orange. The complex depicted is that of the host-guest complex for the oligonucleotide shown with two consecutive **Z**:**P** nucleobase pairs (2P). Within our crystals, the asymmetric unit includes only half of the complex depicted and thus the equivalent of 8 nucleobase pairs and one protein molecule, indicated by the dashed line. Final $2F_o$-F_c electron density maps are shown as blue mesh renderings contoured at 1.0 σ for 8 nucleobase pairs from self-complementary oligonucleotides 2P (B), GC (C), and AT (D). Sequences of GC and AT

oligonucleotides are provided in Table 1. (E) The final $2F_o$-F_c electron density map contoured at 1.4 σ is shown with a stick model for the B7P/G10Z pair in an orthogonal view to that shown in (B).

The 3/6 ZP oligonucleotide did not crystallize in the host-guest system and was subjected to a high-throughput crystallization screen as a complex with the N-terminal fragment of MMLV-RT. Crystals obtained for the oligonucleotide were found to grow without the host protein under conditions including relatively high salt concentrations (~ 2 M ammonium sulfate). As no structural model was available for molecular replacement phasing of the 3/6 ZP structure, oligonucleotides including 5-bromouracil in place of thymine were synthesized and crystallized (Table 1) for experimental phasing purposes. The crystals obtained for the brominated oligonucleotides were actually larger and diffracted to higher resolution than the non-brominated crystals. A Br SAD phasing experiment was performed for the 3/6 ZP Br1 crystals producing a 2.25 Å experimental electron density map of excellent quality as shown in Figure 4, and the structure was refined to 1.98 Å.

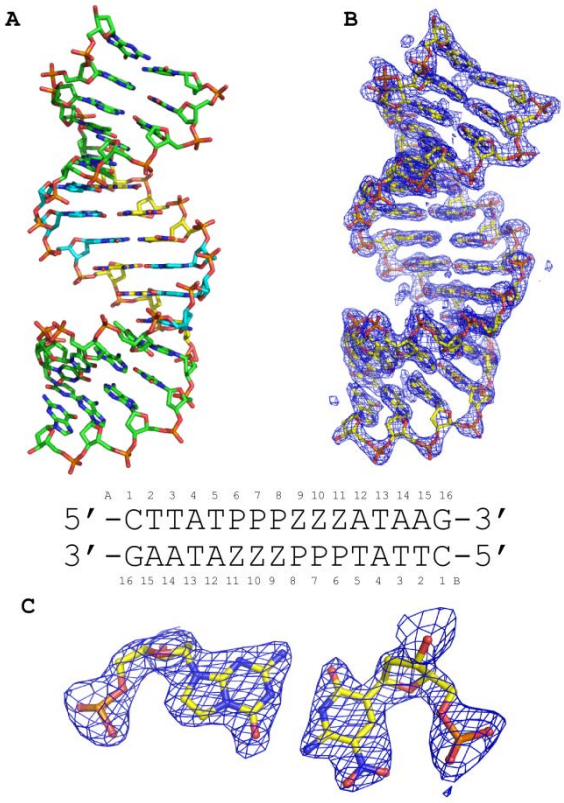

Figure 4. Crystal structure of 16-mer oligonucleotide containing six consecutive **Z**:**P** pairs (3/6 ZP) is shown in (A) as a stick rendering with C, green for non-Z:P pairs, yellow for **Z**, and cyan for **P**, O in red, nitrogen in blue, phosphorous in orange. The 2.25 Å experimental electron density map derived from Br SAD phasing is shown as a blue mesh contoured at 1.5 σ with the entire final refined structural model as a stick model (B) and for the A8P/B9Z pair (C) in an orthogonal view. The sequence of the oligonucleotide is shown along with numbering scheme employed in the coordinate file; T3 in each strand was substituted with 5-bromouracil.

Oligonucleotides analogous to 3/6 ZP containing either A:T or G:C pairs in place of the **Z**:**P** pairs were crystallized as host-guest complexes for comparative analysis with 2P and 3/6ZP sequences (Table 1). Both the AT and GC oligonucleotides crystallized as host-guest complexes, and their structures were determined to 1.68 Å and 1.78 Å, respectively, by molecular replacement. Final $2F_o$-F_c electron density maps are shown in Figure 3 for the DNA in each structure.

While maintaining canonical hydrogen bonding interactions, ZP-containing DNA exhibits inherent flexibility allowing it to adopt both A- or B-form DNA

The literature suggests that the ability of DNA to transition between B- and A-forms is required for it to interact with many proteins, including transcription factors, DNA repair enzymes, and polymerases.[52] Therefore, it was of interest to analyze the 3/6 ZP oligonucleotide that crystallized as A-form DNA as an important biological form of DNA.[53] In this case, one end of the duplex forms hydrogen bonds in the minor groove with **Z**:**P** pairs located in the middle of another DNA duplex, molecular packing interactions often observed in other A-DNA self-standing structures. Specifically, O2 of C1 hydrogen bonds to N2 of P7 and N2 of G16 hydrogen bonds to O2 of Z10 (see Figure 4 for numbering scheme). As analyzed by the software package 3DNA,[54] the helical form of this structure is classified as A-form excluding the first three dinucleotide steps and the 13th dinucleotide step, which includes the 5-bromouracil. All six contiguous **Z**:**P**-containing dinucleotide steps are therefore classified as A-DNA. GC-rich sequences that crystallize in A-form involve similar hydrogen-bonding interactions between the ends and G:C pairs in the middle of the oligonucleotides. The fact that 3/6 ZP contains 8 A:T pairs and yet still crystallized as A-form suggests that **Z**:**P** pairs may have an even higher propensity to adopt A-form DNA than GC-rich sequences. In a survey of A-DNA structures, the oligonucleotide sequences were G:C-rich with most containing fewer than 2 A:T nucleobase pairs. In fact, the structure most similar to 3/6 ZP in terms of length is that of a GC-rich 14 base pair oligonucleotides (PDB ID: 4OKL), which includes only 2 central A:T nucleobase pairs.

In contrast to the 3/6 ZP, 2P, AT, and GC duplexes (Table 1) crystallized in the host-guest system and exhibit B-form throughout, excluding the terminal dinucleotide steps, as analyzed by 3DNA.[54] The AT and GC duplexes are analogous to 3/6 ZP in that the 5'-P_3Z_3 sequence is replaced with 5'-G_3C_3 or 5'-A_3T_3, maintaining the positioning of purine and pyrimidine-like bases within the oligonucleotides. In theory, given that the 3/6 ZP duplex like the AT and GC duplexes adopts B-form

under low salt conditions in solution as assessed by CD, it should have been possible to obtain host-guest crystals. However, negative crystallization results are difficult to interpret, as many factors govern the crystallization of a protein-DNA complex, including solubility, flexibility, and conformational homogeneity of the DNA.

In both B- and A-DNA, N4, N3, and O2 of **Z** hydrogen bond to O6, N1, and N2 of **P**, respectively, with hydrogen bonding distances of 2.7- 3.0 Å typically seen in canonical helical forms of DNA (Figure 5). However, as shown in Figure 5C and D, there is evidence of shearing in both the 6DP:11DZ pair and the 7DP:10DZ pair, with the first pair exhibiting greater shearing (-1.1 Å) than the second (0.8 Å) as analyzed in 3DNA[54]. Shearing of nucleobase pairs in the 2P structure is not unique to the **Z:P** pairs; the central A:T pairs exhibits shearing values of similar magnitude (0.9 Å). **Z:P** pairs from the 3/6 ZP structure exhibit standard geometry and hydrogen bonding distances as shown in Figure 5B with little or no shearing.

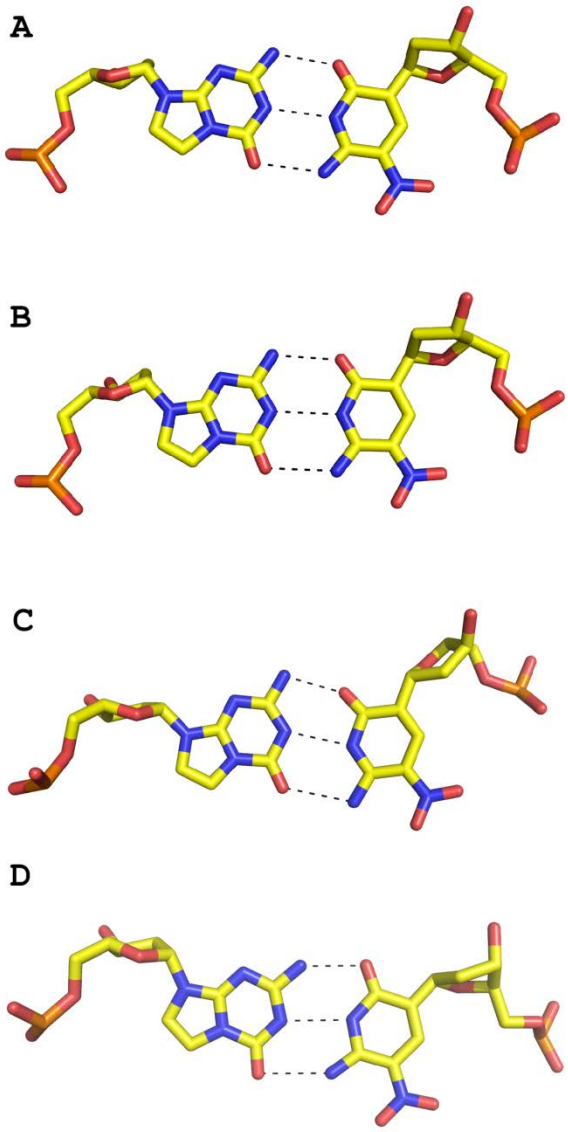

Figure 5. Hydrogen-bonding interactions are shown for the 3/6 ZP structure in (A) for 8 DP (left) and 9 DZ (right) and in (B) 7 DP and 10 DZ (residue names are consistent with the designations in coordinate files). Hydrogen bonding distances from top to bottom are 2.72 and 2.79 Å for N2 to O2, 2.80 and 2.86 Å for N1 to N3, and 2.76 and 2.84 Å for O6 to N4 in P and Z, respectively. Similarly, hydrogen-bonding interactions in the 2P structure are shown in (C) for 6 DP and 11 DZ and in (D) 7 DP and 10 DZ. Hydrogen bonding distances are 2.76 and 2.74 Å for N2 to O2, 2.89 and 2.79 Å for N1 to N3, and 2.96 and 2.74 Å for O6 to N4, in **P** and **Z**, respectively for 6 DP:11 DZ and 7DP:10DZ. **P** and **Z** exhibit shearing in (C) and (D). Refer to Figures 3 and 4 for oligonucleotide numbering schemes.

The Z-nitro group imparts unique structural properties to both A- and B-DNA

Of particular interest are the structural properties associated with the unique NO_2 group in **Z** in the context of A- or B-DNA. Historically, this nitro group was introduced initially to manage the chemical properties of the system presenting a hydrogen bond donor-donor-acceptor pattern (from the major to the minor groove).[26] Subsequently, the nitro group was found to confer binding potential on GACT**ZP** libraries that appears to be absent in standard GACT libraries.[27]

In the A-DNA structure, the **Z** nucleobase, including its NO_2 group, is planar; this appears to facilitate stacking interactions between the NO_2 of one **Z** nucleobase and the pyrimidine (or purine) ring of the adjacent nucleobase (Figure 6 A and B). One oxygen atom of the nitro group in these stacked planar **Z** nucleobases is within hydrogen-bonding distance (~2.7 Å) of N4 of the pyrimidine-like ring. In contrast, in B-DNA, the NO_2 group does not stack with adjacent bases but is nearly planar in both **Z** nucleobases with similar hydrogen-bonding between a nitro oxygen and N4 to that observed in 3/6ZP (Figure 6 C and D). Accordingly, the electron density for the **Z**-NO_2 groups in B-DNA (Figure 3E) is not as well ordered as in A-DNA (Figure 4C), consistent with the more constrained conformation observed in the stacking interactions in A-form DNA. Potentially, the favorable stacking interactions of the Z-NO_2 groups contribute to the ability of the 16-mer oligonucleotide including only six **Z**:**P** pairs to crystallize as A-form DNA. The fact that **Z** is a C-glycoside with a carbon-carbon linkage to its deoxyribose sugar rather than nitrogen-carbon linkage found in natural bases may also be a contributing factor.

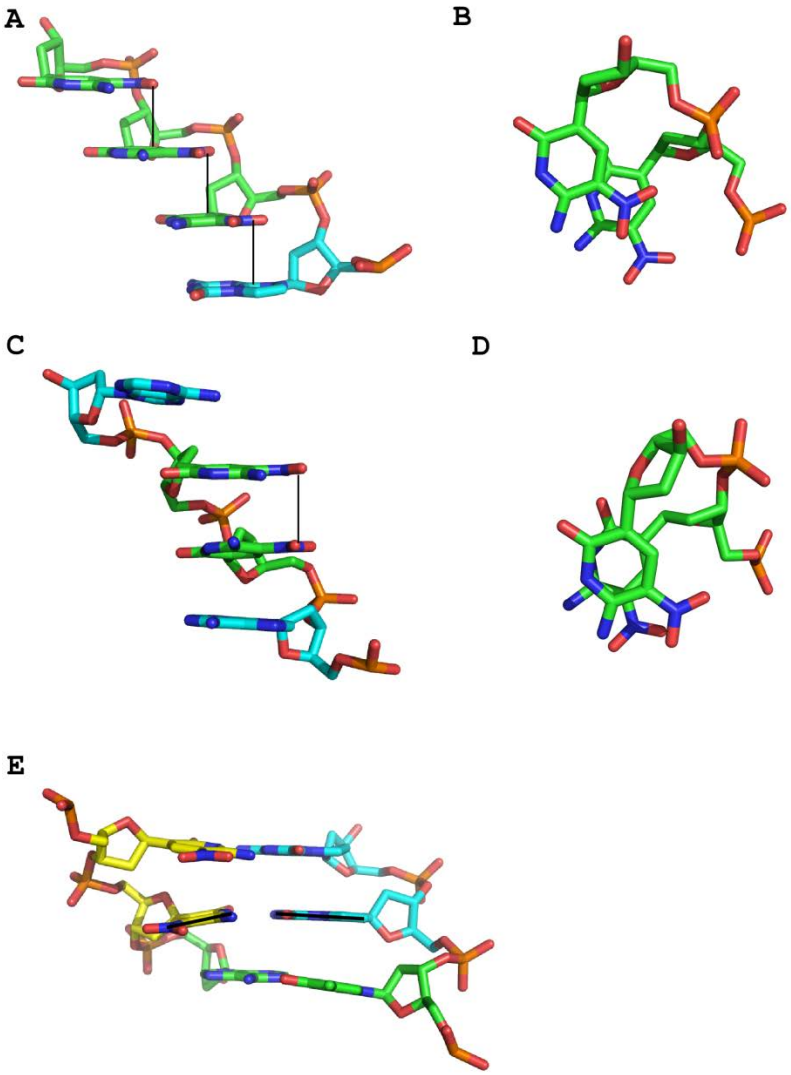

Figure 6. Structural characteristics of **Z**:**P** nucleobase pairs. (A) In A-DNA, the **Z**-NO_2 stack over the pyrimidine or purine ring of the adjacent nucleobase emphasized by solid lines as shown in this stick renderings C, green for **Z**, cyan for P, O in red, N in blue, and phosphorous in orange for $P_8Z_9Z_{10}Z_{11}$, A chain. (B) Orthogonal view shown for two **Z:P** pairs. (C) In B-DNA, the **Z**-NO_2 groups are positioned roughly over one another but do not stack with nucleobase rings. Rather, the pyrimidine rings stack with adjacent nucleobase rings as shown in this stick rendering for bases for $A_9Z_{10}Z_{11}A_{12}$ for the G chain. (D) Orthogonal view shown for the two consecutive **Z:P** pairs in B-DNA. (E) The central **Z**:**P** pair, with a black line indicating the trajectory of each nucleobase, exhibits a large buckle angle as shown in this

stick rendering for T_5**P**$_6$**P**$_7$ (B chain) and **Z**$_{10}$**Z**$_{11}$A$_{12}$ (G chain). C atoms are indicated in yellow for **Z**, cyan for **P**, and green for A or T, with other atoms as indicated above.

To characterize the DNA parameters of **Z**:**P** pairs in A-DNA, we analyzed the local base pair step parameters, local base pair helical parameters, and the groove widths for the central **Z**:**P** pairs in our structure and compared them to those of the GC-rich 4OKL structure (5'-CCCCGGTACCGGGG) using the program 3DNA[54]. Average values were calculated for five **Z**:**P** or **P**:**Z** dinucleotide steps for our structure and 10 GC or CG dinucleotide steps from 4OKL. As previously noted, A:T pairs are found infrequently in A-DNA structures and thus are not available for comparison. As shown in Table 3, the average rise, roll, and twist values for **Z**:**P** dinucleotide steps are similar to the values observed for GC dinucleotide steps and the average parameters for A-DNA reported by Lu et al.[54] However, the average slide value for **Z**:**P** pairs of -2.1 Å in A-DNA is slightly larger than observed in G:C pairs in 4OKL (-1.9 Å) and on average in A-DNA structures -1.5 +/- 0.3 Å. This finding is consistent with the preferential stacking of the NO_2 group with the adjacent pyrimidine or purine ring.

Table 3: DNA parameters

	A-ZP	**A-GC**	**B-ZP**	**B-GC**	**B-AT**
Local bp step					
Slide (Å)	-2.1 (0.2)*	-1.9 (0.4)*	-0.05 (0.6)	0.3 (0.8)	-0.4 (0.1)
Rise (Å)	3.2 (0.05)	3.4 (0.2)	3.3 (0.3)	3.3 (0.2)	3.3 (0.1)
Roll (º)	4.3 (2.5)	3.6 (6.0)*	2.3 (5.9)	0.4 (6.2)	-2.9 (3.9)
Twist (º)	29.5 (1.8)	29.2 (4.0)	30.0 (6.0)	32.8 (11.3)	36.4 (2.9)

Local bp helical					
H-rise (Å)	2.9 (0.2)	3.2 (0.3)	3.3 (0.2)	3.3 (0.2)	3.3 (0.1)
Inclination (º)	8.1 (4.7)	6.3 (10.2)*	3.1 (11.9)	3.6 (13.0)	-4.5 (6.5)
H-twist (º)	29.8 (2.0)	30.0 (4.8)	31.0 (6.3)	33.4 (11.1)	36.7 (2.8)
X-displacement (Å)	-4.8 (0.4)	-4.5 (0.8)	-0.8 (0.7)	-0.09 (1.7)	-0.2 (0.6)
Groove width					
Major (Å)	18.9 (1.2)	18.0 (1.6)	18.7 (0.9)	18.0 (0.6)	19.1 (0.6)
Minor (Å)	16.5 (0.6)	16.9 (0.4)	12.7 (0.6)	12.4 (0.9)	9.7 (0.5)

Dinucleotide steps 5-11 for A-ZP and dinucleotide steps 1-6 and 8-13 for A-GC were analyzed from our structure and 4OKL, respectively. For B-ZP, B-GC, and B-AT, dinucleotide steps 5-7 were analyzed from our host-guest structures. Standard deviations are indicated in parentheses. (*) denotes values that are outside the average range reported by Lu et al. [54]

Otherwise, local base pair helical parameters including H-rise, inclination, H-twist, and X-displacement are all within the average range observed in A-DNA structures and are similar to those observed for G:C pairs in 4OKL. Within the central **Z**:**P** pairs of our structure, the average major groove width of 18.9 Å is approximately 1 Å wider the groove width associated with the G:C pairs (18.0 Å) in the 4OKL structure. Widening of the major groove may result from the presence of the **Z**-NO_2 in the major groove. The minor groove widths for both **Z**:**P** and G:C pairs in A-DNA are more similar than the major groove widths with average values of 16.5 and 16.9 Å, respectively.

Similarly, the properties of the **Z**:**P** nucleobase pairs in B-DNA were assessed and compared to A:T and G:C pairs at the same positions in the DNA sequence in our host-guest structures. There are

four **Z:P** pairs within six nucleotides (5'-**PP**TA**ZZ**) in the 16-mer sequence crystallized; however, only half of this sequence and its complementary sequence are unique as the other half is related by crystallographic symmetry. These sequences are referred to as B-**ZP**, B-GC, and B-AT in contrast to those in the A-DNA structures, which are A-**ZP** and A-GC. In this case, the three dinucleotide steps containing **ZP** pairs and the analogous dinucleotide steps from the AT and GC oligonucleotide duplex structures were analyzed. The average local base pair step parameters and local base pair helical parameters for B-**ZP** all fall within the range observed for B-DNA structures. Average values for each of these parameters for B-**ZP** are more similar to those calculated for B-GC than for B-AT. As was true for A-**ZP**, the major groove width of B-**ZP** of 18.7 Å is significantly (~ 0.7 Å) wider than that of B-GC (18.0 Å). However, it is slightly narrower than B-AT, which has an average major groove width of 19.1 Å. The minor groove width for B-**ZP** of 12.7 Å is similar to that of B-GC (12.4 Å) and significantly wider than that of B-AT (9.7 Å). AT-rich sequences are known to exhibit deep, narrow minor grooves, which bind to a number of small molecules.

Local base pair parameters including shear, stretch, stagger, buckle, propeller, and opening were also analyzed using 3DNA[54] for A-**ZP** pairs and found to be fairly uniform in value. Propeller values of -10.8 º and -13.7 º are the largest observed values and occur at the junctions between the **PZ** or **ZP** and natural nucleobase pairs. For B-**ZP**, the pair Z_6P_{11} exhibits an unusually large buckle angle of -14.2 º (Figure 6 E), which is much larger than buckle angles of -1.4 and 5.4 º exhibited for equivalent A:T or G:C base pairs, respectively. This large buckle angle for Z_6P_{11} may contribute to the shearing observed for this base pair. In the major groove, the **Z**:**P** pair provides novel major groove hydrogen bonding opportunities with four electronegative atoms, two provided by the NO_2 group, as compared to three present in either A:T or G:C pairs (Figure 7 A-C). In the minor groove, **Z**:**P** pairs present the same hydrogen bonding pattern of electronegative atoms as the GC base pair (Figure 7 D-F), with all three nucleobase pairs including O2 and N3 atoms.

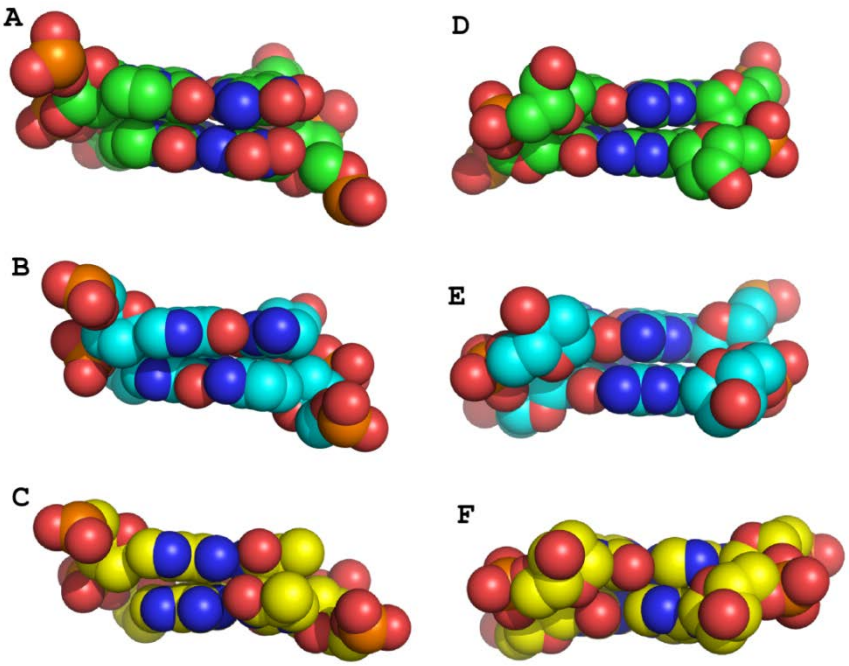

Figure 7. Major and minor groove views of **Z**:**P**, G:C, and A:T pairs in B-DNA. Dinucleotides for $\mathbf{Z}_{10}\mathbf{Z}_{11}/\mathbf{P}_{6}\mathbf{P}_{7}$ (A) and the equivalent pairs for GC (B) and AT (C) host-guest complexes are shown as Van der Waals sphere renderings for the major groove with C in green for **Z**:**P**, cyan for GC, and yellow for AT. Minor groove views are shown for **Z**:**P** (D), GC (E), and AT (F).

Thus, these studies show that **Z**:**P** nucleobase pairs are accommodated in multiple and consecutive positions in both A- and B-DNA, with the DNA parameters quite similar to those observed for G:C pairs in the same helical form. This is despite the unnatural nitro group carried by **Z**. The most significant impact of **Z**:**P** pairs on the structure of the DNA duplexes is widening of the major groove in both A and B-DNA as compared to values calculated for similar GC regions. We might speculate that the NO_2 group, which faces the major groove, creates this effect.

Absent experimental data, it is difficult to predict what sequences favor the A-form versus B-form DNA. It is conceivable that multiple **Z**:**P** nucleobase pairs have a higher propensity to form A-DNA than equivalent sequences with G:C pairs. This is consistent with our failure to obtain crystals of

the 3/6 ZP sequence in our host-guest system, which selects for B-form oligonucleotides. It is also consistent with the observation of the A-form duplex in 3/6 ZP sequence; other examples of A-DNA are more G:C-rich. However, the CD spectrum of the 3/6 ZP oligonucleotide in a low salt buffering solution has features consistent with B-DNA and is similar overall to that obtained for the equivalent GC sequence. This suggests that at least in 3/6 ZP, both the A- and B-forms are accessible upon changing salt concentrations.

Conclusions

Expansion of genetic alphabets that combine standard and non-natural nucleobase pairs in duplex DNA will be an essential element in the evolution of efforts in synthetic biology. The need to maintain (or at least prevent radical divergence from) existing canonical structures, however, represents an important constraint upon the incorporation of synthetic nucleotides into DNA if existing nucleic acid binding proteins or enzymes are to act on them. Of course, even canonical structures have considerable diversity, and the special properties of stacked A:T versus G:C nucleobase pairs in duplex DNA are well documented.[47] Thus, characterizing the structural properties of DNA containing multiple and/or contiguous non-natural nucleobase pairs as compared to natural nucleobase pairs has considerable importance.

Three important findings emerge from our structural studies. The first is the discovery that canonical Watson-Crick pairing survives in duplexes that contain multiple and multiple adjacent **Z:P** pairs. This is found both with a host and without a host, evidence that this geometry is intrinsic to the **Z:P** pair. The second is that the **ZP**-containing oligonucleotides adopt canonical helical forms, B- and A-form DNA. The ability of DNA to adopt A-form enables a number of important protein-DNA interactions such as those in the polymerase active site, which requires that the nucleobase pair within the active site and the adjacent pair adopt A-form in order to appropriately position the template-primer

for optimal interactions with the polymerase increasing fidelity.[52] The third is that the Z-nitro group imparts new properties to the major groove of DNA that can potentially be exploited for recognition by proteins.

These results have implications for the thread of synthetic biology that adds nucleotides to standard DNA creating expanded DNA to be used by polymerases, ligases, kinases and other enzymes, its use as a platform for *in vitro* evolution to create functional molecules, and ultimately its functioning inside of living cells. The nitro group, perhaps a "universal" binding moiety, can be incorporated in multiple sites without disrupting the duplex geometry. This allows natural polymerases to search effectively the sequence space made possible by added nucleotide "letters". Last, these results show that ZP-containing oligonucleotides have the conformational plasticity of natural DNA, a plasticity that may be necessary for their function in living cells.

ACKNOWLEDGEMENTS

This work was partially supported by grants R01DK061666 (NGJR) and R01GM111386 (SAB) from the National Institutes of Health, HDTRA1-13-1-0004 from the Defense Threat Reduction Agency, NNX14AK37G from the National Aeronautics and Space Administration (SAB), and the Templeton World Charitable Foundation (SAB). Data for this work were collected at the Advanced Photon Source at SBC, 19-ID and 19-BM beamlines and GM/CA, 23-ID-D beamline, which has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357. We thank Z. Otwinowski and W. Minor for helpful discussions while at the SBC beamline and Marianne Cuff and Rushlan Sanishvili for assistance during data collection at SBC and GM/CA, respectively.

REFERENCES

(1) Leduc, S. *La biologie synthétique*; A. Poinat: Paris, 1912.

(2) Szybalski, W. In *Control of Gene Expression*; Kohn, A., Shatkay, A., Eds.; Plenum Press: New York, 1974.

(3) Sismour, A. M.; Benner, S. A.; *Expert Opin Biol Ther* **2005**, *5*, 1409.

(4) Rappaport, H. P.; *Nucleic Acids Res* **1988**, *16*, 7253.

(5) Switzer, C.; Moroney, S. E.; Benner, S. A.; *J Am Chem Soc* **1989**, *111*, 8322.

(6) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A.; *Nature* **1990**, *343*, 33.

(7) Ishikawa, M.; Hirao, I.; Yokoyama, S.; *Tetrahedron Letters* **2000**, *41*, 3931.

(8) Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E.; *J Am Chem Soc* **2001**, *123*, 7439.

(9) Kool, E. T.; *Acc Chem Res* **2002**, *35*, 936.

(10) Geyer, C. R.; Battersby, T. R.; Benner, S. A.; *Structure* **2003**, *11*, 1485.

(11) Henry, A. A.; Romesberg, F. E.; *Curr Opin Chem Biol* **2003**, *7*, 727.

(12) Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A.; *J Am Chem Soc* **2003**, *125*, 9970.

(13) Benner, S. A.; *Acc Chem Res* **2004**, *37*, 784.

(14) Hirao, I.; Harada, Y.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Yokoyama, S.; *J Am Chem Soc* **2004**, *126*, 13298.

(15) Leal, N. A.; Kim, H. J.; Hoshika, S.; Kim, M. J.; Carrigan, M. A.; Benner, S. A.; *ACS Synth Biol* **2015**, *4*, 407.

(16) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A.; *Nature* **1992**, *356*, 537.

(17) Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S.; *Nat Methods* **2006**, *3*, 729.

(18) Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Nishikawa, K.; Yokoyama, S.; *Nat Biotechnol* **2002**, *20*, 177.

(19) Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E.; *Proc Natl Acad Sci U S A* **2012**, *109*, 12005.

(20) Malyshev, D. A.; Pfaff, D. A.; Ippoliti, S. I.; Hwang, G. T.; Dwyer, T. J.; Romesberg, F. E.; *Chemistry* **2010**, *16*, 12650.

(21) Malyshev, D. A.; Seo, Y. J.; Ordoukhanian, P.; Romesberg, F. E.; *J Am Chem Soc* **2009**, *131*, 14620.

(22) Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E.; *J Am Chem Soc* **2009**, *131*, 3246.

(23) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correa, I. R., Jr.; Romesberg, F. E.; *Nature* **2014**, *509*, 385.

(24) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F. E.; Marx, A.; *J Am Chem Soc* **2013**, *135*, 18637.

(25) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Dwyer, T. J.; Ordoukhanian, P.; Romesberg, F. E.; Marx, A.; *Nat Chem Biol* **2012**, *8*, 612.

(26) Hutter, D.; Benner, S. A.; *J Org Chem* **2003**, *68*, 9839.

(27) Sefah, K.; Yang, Z.; Bradley, K. M.; Hoshika, S.; Jimenez, E.; Zhang, L.; Zhu, G.; Shanker, S.; Yu, F.; Turek, D.; Tan, W.; Benner, S. A.; *Proc Natl Acad Sci U S A* **2014**, *111*, 1449.

(28) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A.; *J Am Chem Soc* **2011**, *133*, 15105.

(29) Yang, Z.; Chen, F.; Chamberlin, S. G.; Benner, S. A.; *Angew Chem Int Ed Engl* **2010**, *49*, 177.

(30) Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A. M.; Benner, S. A.; *Nucleic Acids Res* **2006**, *34*, 6095.

(31) Sun, D.; Jessen, S.; Liu, C.; Liu, X.; Najmudin, S.; Georgiadis, M. M.; *Protein Sci* **1998**, *7*, 1575.

(32) Otwinowski, Z.; Minor, W.; *Methods Enzymol.* **1997**, *276*, 307.

(33) Schneider, T. R.; Sheldrick, G. M.; *Acta Crystallogr D Biol Crystallogr* **2002**, *58*, 1772.

(34) Sheldrick, G. M. In *Zeitschrift für Kristallographie/International journal for structural, physical, and chemical aspects of crystalline materials* 2002; Vol. 217, p 644.

(35) Cowtan, K.; *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 470.

(36) Cowtan, K.; Main, P.; *Acta Crystallogr D Biol Crystallogr* **1998**, *54*, 487.

(37) Cowtan, K. D.; Zhang, K. Y.; *Prog Biophys Mol Biol* **1999**, *72*, 245.

(38) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S.; *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 235.

(39) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J.; *Acta Crystallogr D Biol Crystallogr* **1997**, *53*, 240.

(40) Humphrey, W.; Dalke, A.; Schulten, K.; *J Mol Graph* **1996**, *14*, 33.

(41) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A.; *J Comput Chem* **1993**, *14*, 1347.

(42) Ditchfield, R.; Hehre, W. J.; Pople, J. A.; *The Journal of Chemical Physics* **1971**, *54*, 724.

(43) Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H.; *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213.

(44) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K.; *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486.

(45) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J.; *J Appl Crystallogr* **2007**, *40*, 658.

(46) *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 760.

(47) Neidle, S. *Nucleic Acid Structure*; Oxford University Press: New York, 1999.

(48) Singh, I.; Jian, Y.; Li, L.; Georgiadis, M. M.; *Acta Crystallogr D Biol Crystallogr* **2014**, *70*, 752.

(49) Goodwin, K. D.; Lewis, M. A.; Long, E. C.; Georgiadis, M. M.; *Proc Natl Acad Sci U S A* **2008**, *105*, 5052.

(50) Crowther, R. L.; Remeta, D. P.; Minetti, C. A.; Das, D.; Montano, S. P.; Georgiadis, M. M.; *Proteins* **2004**, *57*, 15.

(51) Montano, S. P.; Cote, M. L.; Roth, M. J.; Georgiadis, M. M.; *Nucleic Acids Res* **2006**, *34*, 5353.

(52) Lu, X. J.; Shakked, Z.; Olson, W. K.; *J Mol Biol* **2000**, *300*, 819.

(53) Wahl, M. C.; Sundaralingam, M.; *Biopolymers* **1997**, *44*, 45.

(54) Lu, X. J.; Olson, W. K.; *Nucleic Acids Res* **2003**, *31*, 5108.

TOC Graphic